

Δαίμων. Revista de Filosofía, n° 34, 2005, 97-110

## El problema mente-cuerpo reconsiderado<sup>1</sup>

DIANA I. PÉREZ\*

**Resumen:** En este trabajo me propongo reconsiderar ciertos argumentos centrales acerca del llamado «problema mente-cuerpo», partiendo de una peculiar manera de interpretar a los conceptos mentales ordinarios: como conceptos de clase natural. En la primera parte desarrollo esta propuesta y en la segunda examino los argumentos de Kripke en contra de la teoría de la identidad psicofísica, el argumento de los zombis en contra del funcionalismo y el argumento de Churchland a favor del eliminativismo. Concluyo esbozando una respuesta al problema mente-cuerpo a mitad de camino entre el funcionalismo y la teoría de la identidad.

**Palabras clave:** problema mente-cuerpo; conceptos mentales; *Folk Psychology*; funcionalismo; teoría de la identidad; clases naturales.

**Abstract:** In this paper I shall offer a reconsideration of three main arguments in the current debate on the mind-body problem, on the light of a peculiar way of conceiving mental concepts: I shall defend the view that mental concepts have to be considered as natural kind concepts. In the first part, I shall develop this proposal and in the second part I shall examine Kripke's arguments against the identity theory, the zombi's argument against functionalism and Churchland's argument for eliminativism. I conclude the paper with an sketchy answer to the mind-body problem, in the middle of the road between functionalism and the identity theory.

**Key words:** Mind-body problem, mental concepts, Folk psychology, functionalism, identity theory, natural kinds.

«Somos, cuando filosofamos, como salvajes, hombres primitivos, que oyen los modos de expresión de los hombres civilizados, los malinterpretan y luego extraen las más extrañas conclusiones de su interpretación». (Wittgenstein 1953, §194)

Mucho se ha escrito, especialmente durante el siglo XX, alrededor del «problema mente-cuerpo». La variedad de respuestas examinadas para este problema parecen definidas: dualismo (de sustancias o de propiedades), conductismo (psicológico o lingüístico), reduccionismo fisicalista, anomalismo, funcionalismo (analítico o psicofuncionalismo), eliminativismo. Sin embargo, estas

Fecha de recepción: 13 febrero 2004. Fecha de aceptación: 28 octubre 2004.

\* Departamento de Filosofía. Universidad de Buenos Aires. Dirección laboral: Bulnes 642 (1176) Ciudad Autónoma de Buenos Aires, Argentina. E.mail: [dperez@filo.uba.ar](mailto:dperez@filo.uba.ar)

1 He leído versiones anteriores de este trabajo en la Universidad de las Islas Baleares, en la Universidad de Murcia y en la Universidad Carlos III (Madrid). Quiero agradecer los valiosos aportes de las diversas audiencias, los comentarios puntuales de G. Fernández Díez, las detalladas y profundas observaciones críticas realizadas por J. Corbí, y los comentarios enriquecedores de J. Vega y F. Broncano.

discusiones, generalmente, dejan de lado o dan por sentado dos cuestiones, que, a mi juicio, deberían examinarse más detenidamente. En primer lugar, muchas de estas discusiones dejan en suspenso la caracterización de «lo mental» y «lo físico» que están a los lados de la relación que se pretende elucidar. Se toman decisiones rápidas y más o menos arbitrarias acerca de qué entender por cada uno de los términos involucrados en la relación psicofísica: así, se promulga a la intencionalidad, o a la conciencia, o a la subjetividad como «el» rasgo de lo mental, y a «lo que la física ideal diga» como caracterizador de lo físico, y sin más reparos comienzan los argumentos acerca de la relación mente-cuerpo. No me voy a detener en esta ocasión en este punto, aunque creo que merece cierta atención, y que no debería despacharse tan rápidamente de las discusiones como suele hacerse.

En este trabajo querría detenerme en otro tema que se da por supuesto también en estas discusiones, y que a mi juicio requiere de un exámen más detallado del que se le suele dar. Es el problema de la naturaleza de nuestros conceptos mentales ordinarios, esto es de qué significan términos como «dolor», «temor», «creencia», «deseo», «intención», «asco», «alegría», etc. En este trabajo, en primer lugar, ofreceré una manera específica de entender a estos conceptos: los asimilaré a la categoría de «concepto de clase natural» y desarrollaré el sentido de esta propuesta. En segundo lugar, mostraré cómo varios argumentos centrales pro o contra diversas «posiciones mente-cuerpo» mencionadas arriba suponen una comprensión diferente a la mía de los conceptos mentales, e intentaré mostrar que muchos argumentos clásicos podrían refutarse enteniendo a los conceptos mentales tal y como yo propongo. Mi intención, a pesar de la cita de Wittgenstein que encabeza el artículo, no es *disolver* el problema mente cuerpo, sino reexaminar ciertos argumentos influyentes de la filosofía contemporánea a la luz de una reconsideración de los conceptos mentales que nos permitirá orientarnos hacia una respuesta más adecuada al problema mente-cuerpo.

## Los conceptos mentales como conceptos de clase natural

De acuerdo con la ya clásica propuesta semántica de Putnam (1975) para los términos de clase natural<sup>2</sup>, el significado de estos términos es una n-tupla constituida por marcadores sintácticos, marcadores semánticos, un estereotipo, y un elemento indexical que apunta a la extensión del término. Dejando de lado los dos primeros elementos<sup>3</sup>, la propuesta de Putnam tiende a incorporar en el significado de estos términos dos elementos que las teorías semánticas tradicionales habían dejado de lado: la contribución del mundo real, y la contribución de la sociedad. (p. 50)

- 
- 2 Putnam cree que esta teoría del significado es aplicable a más tipos de términos, incluyendo artefactos, etc. Yo aplicaré las consideraciones de Putnam acerca del significado de «significado» a los conceptos mentales exclusivamente. Tal vez, la idea de interpretarlos como conceptos de clase natural resulte osada pero, como espero que quede claro más adelante, mi propósito central al proponer esta interpretación es la de remarcar que los conceptos mentales poseen una «esencia» escondida (en un sentido que aclararé) accesible al estudio científico (en un sentido muy amplio del término), tal como Putnam muestra en su trabajo que ocurre con el concepto de agua. Empezar con la propuesta de Putnam es un recurso expositivo, como se verá más adelante me aparto en muchos aspectos de su propuesta. También se me ha objetado la interpretación de los conceptos mentalistas como conceptos de clase natural, señalándoseme que el adjetivo «natural» queda desdibujado en mi propuesta, al aceptar una comprensión de lo científico tan amplia como la que yo propongo, que incluye sociología, antropología, etc. e incluso he llegado a admitir, en conversaciones con J. Corbí y algunos colegas de la Universidad de Murcia, conocimientos nada científicos como los provenientes de ficciones literarias, aunque sólo de manera provisoria, en aquellos campos donde aun no ha sido desarrollada una adecuada teoría científica. Tal vez la expresión «concepto de clase natural» no sea del todo apropiada para mi propuesta, pero no se me ocurre ninguna mejor. Espero que mis ideas queden claras en el trabajo, invito al lector a dejar de lado prejuicios que provienen de la carga semántica de este expresión, y a hacer un esfuerzo por comprender mis ideas desprejuiciadamente.
  - 3 El marcador sináctico, por ejemplo «sustantivo» y el marcador semántico, por ejemplo «líquido», «animal», etc.

La contribución del mundo al significado de los términos de clase natural está encarnada en el componente indexical implícito en todo término de clase natural. En el caso de «agua», su componente indexical apunta a una sustancia en el mundo cuya descripción correcta ha sido descubierta por la química, se trata de la sustancia cuya descripción científica es «H<sub>2</sub>O». El argumento de Putnam para mostrar que el mundo debe ser incorporado al significado del término «agua» es el bien conocido argumento de las Tierras Gemelas que no voy a desarrollar aquí. El componente indexical del significado es el que permite incorporar el mundo al significado, y esta idea es combinada por Putnam con una propuesta metafísica realista. Así sostiene: «En la visión que propugno, cuando Arquímedes afirmaba que algo era oro, no solamente decía que tenía las características superficiales del oro... lo que decía era que tenía la misma *estructura general oculta* (la misma «esencia» por así decirlo) que cualquier pieza normal de oro local». (p. 36) Sin embargo, a pesar de este compromiso realista explícito, Putnam se da cuenta de que si bien todos los términos de clase natural apuntan hacia una estructura oculta, el mundo no colabora de igual manera en todos los casos: así, en algunos casos la ciencia descubrirá efectivamente la estructura oculta hacia la que el término ordinario apuntaba, como por ej. H<sub>2</sub>O en el caso de «agua», pero en otros casos puede ocurrir que los miembros de la extensión de una clase dada no tengan finalmente ninguna estructura oculta en común, como ciertas enfermedades, que podrían tener en común finalmente sólo un conjunto de síntomas; en otros casos podría ocurrir que la estructura oculta fuera sólo una etiología (Putnam ejemplifica este caso con la tuberculosis); o podría ocurrir que hubiera dos estructuras ocultas allí donde se esperaba una, como en el caso de «jade». El punto de Putnam aquí no es completamente claro: termina afirmando un condicional: «si existe una estructura oculta, entonces ésta determinará qué es ser miembro de la clase natural» (p. 44) pero no indica qué ocurre con los casos en los que finalmente no se descubre ninguna estructura oculta, o se descubre más de una.

Por otro lado, la contribución de la sociedad al significado de los términos se manifiesta a través de la presencia del estereotipo en el «vector» significado<sup>4</sup>, Putnam dedica amplios pasajes del artículo a elucidar su noción de «estereotipo». El estereotipo es aquello que puede agregarse al marcador semántico, y usualmente suele hacerse, a la hora de dar una descripción del significado de un término de clase natural. Según Putnam «los rasgos centrales del estereotipo son *criterios*: rasgos que en situaciones normales constituyen formas de reconocer si una cosa pertenece a la clase o por lo menos, condiciones necesarias (o condiciones necesarias probabilísticas) para la membresía en la clase». (p. 27) Aprender el estereotipo es aprender la palabra, porque, de acuerdo con la hipótesis de la división de la tarea lingüística que él defiende, no es necesario aprender a reconocer qué es oro (o sea si una cierta porción de metal forma parte efectivamente o no de la extensión del término «oro») para aprender las palabras con fines comunicativos. En este sentido, sostiene lo siguiente: «Diremos que alguien *adquirió* la palabra tigre si es capaz de usarla de tal forma que 1) su uso llena los requisitos (es decir, la gente no le dice cosas tales como «no sabe qué es un tigre» ...) y 2) su manera de situarse en el mundo y en su comunidad lingüística es tal que, en su idiolecto, la extensión socialmente determinada de la palabra «tigre» es el conjunto de los tigres». (p. 53) Y continúa: «La comunicación significativa requiere que la gente sepa algo sobre lo que está hablando. ... Lo que sostengo es que se *requiere* que los hablantes sepan algo de los tigres (estereotípicos) para que pueda consi-

4 También se manifiesta a través de la tesis de la división de la tarea lingüística y la existencia de una comunidad de expertos que colaboran en la redesccripción científica del elemento indexical. Agradezco a J. Vergara hacerme reflexionar sobre este punto. Sin embargo, estoy interesada aquí en subrayar el rol de la sociedad en la determinación de los rasgos estereotípicos del concepto, de las relaciones criterios en las que está inmerso.

derarse que adquirieron la palabra «tigre». ... La comunidad lingüística también tiene sus patrones mínimos con respecto a su sintaxis y a su «semántica». ... Sin embargo, la naturaleza del nivel mínimo de competencia depende fuertemente de la cultura y del tema. La comunidad lingüística exige a los hablantes del inglés que sean capaces de distinguir los tigres de los leopardos; no les exige que sean capaces de distinguir los olmos de las hayas». (p. 55)

Resumiendo, Putnam sostiene que los estereotipos indican cuál es el aspecto de un X o cómo actúa o cómo es<sup>5</sup>. Por ej, para que alguien haya adquirido «tigre» debe saber que los tigres tienen rayas. Pero también sostiene que esto no es una verdad analítica (p. 57). En efecto, podría haber casos de tigres sin rayas: un tigre albino no es una entidad contradictoria. La información del estereotipo podría ser incorrecta, esto se explica por el hecho de que los estereotipos captan rasgos pertenecientes a los miembros paradigmáticos de la clase en cuestión. (p. 58) (es *a priori*, pero no es necesaria)

Como dije más arriba, el componente indéxico de los términos de clase natural señalan una estructura oculta, una esencia, que colabora en la constitución del significado del término. Sin embargo, esta participación de una «esencia» en la constitución del significado de los términos de clase natural, se puede entender de dos maneras diferentes. Por un lado, se puede dar de esta tesis una versión esencialista metafísica, una lectura realista, de acuerdo con la cual el significado del término está en parte constituido por una esencia que tiene existencia independiente del hablante, cuya naturaleza eventualmente el hablante puede descubrir por medios empíricos. Esta es la lectura habitual que se ha hecho de la propuesta de Putnam. Sin embargo, esta lectura tiene ciertas complicaciones que se originan en el hecho de que, como el mismo Putnam reconoce, el mundo no colabora siempre de la misma manera, y en algunos casos puede llegar a ser inexistente la «esencia» preteóricamente involucrada en un término dado, o puede ser múltiple, etc. Como señalamos más arriba, en estos casos, no está claro qué pasa con el significado del término de clase natural: si un término dado señalaba la existencia de una esencia única, y esta no se encuentra, ¿hay que abandonar nuestro lenguaje ordinario? Si se descubren más esencias, ¿hay que distinguir dos términos diferentes con distinta referencia pero idéntico estereotipo?

Una manera, a mi entender más adecuada, de entender el funcionamiento del componente indéxico en el significado de estos términos, es suponer que lo que contribuye al significado de los términos de clase natural no es una esencia en el mundo, sino un cierto esencialismo a nivel psicológico. De acuerdo con esta propuesta, los términos de clase natural funcionan postulando la existencia de mecanismos ocultos, lo que lleva a los individuos a postular nuevas teorías con el fin de poder dar cuenta de tales mecanismos. El componente indéxico señala una incompletud del término en cuestión que será completado con el avance del conocimiento, con la postulación de nuevos mecanismos explicativos de los fenómenos involucrados. Como se puede apreciar, esta lectura del componente indéxico es perfectamente compatible con una posición que no sea realista respecto de las teorías postuladas<sup>6</sup>. La idea que está detrás de esta propuesta podría resumirse en los siguientes términos: los términos de clase natural tienen un significado incompleto, pero la manera de completarlos no es incorporando el mundo en el significado, sino con más teoría, con más actividad cognitiva por parte del sujeto. No hay por qué comprometerse con el realismo metafísico para defender la

5 Nótese que Putnam no afirma que los estereotipos de las clases naturales estén constituidos por las propiedades fenomenológicas de las cosas, ni por las propiedades funcionales exclusivamente. Están constituidos por aquellos rasgos que resultan relevantes para el aprendizaje de los términos en cuestión en nuestra vida cotidiana, sean los que fueren.

6 Esta es la lectura que hacen los psicólogos cognitivos, defensores de la visión teórica de los conceptos, cf. Murphy y Medin (1985), Medin (1989).

idea de que detrás de los términos de clase natural hay la postulación de un esencialismo psicológico, que lleva a los individuos a desarrollar nuevas teorías, más abarcativas de ciertos fenómenos cotidianos. En realidad, esta manera de entender los términos de clase natural permite recoger el espíritu de la propuesta quineana acerca de las clases naturales, quien sostiene que a partir de cierto momento del desarrollo, el niño deja de categorizar el mundo considerando exclusivamente las semejanzas superficiales y comienza a categorizarlo a la luz de teorías más explicativas. Esta idea ha sido además testada empíricamente y ha dado lugar a lo que se conoce dentro del ámbito de la psicología como «visión teórica o visión explicativa» de los conceptos<sup>7</sup>. Y como es sabido el esencialismo de cualquier tipo está bien alejado del espíritu de Quine.

En cuanto al componente «estereotípico» de los conceptos de clase natural, propongo aclarar la noción de «estereotipo» que tengo en mente y que creo que no es muy diferente de la que Putnam tenía en mente, (aunque argumentar esto aquí excede los límites de este trabajo), apelando a la noción de «criterio» Wittgensteiniana, a partir de la presentación de la misma que hace Hacker (1972) Cap. X. La idea es que hay un tipo peculiar de relación entre oraciones que contribuye a determinar el significado de los términos que ocurren en dichas oraciones. Esta relación, denominada «criterial», no es una relación de condiciones necesarias y suficientes, ni es una relación de implicación lógica (*entailment*) ni es una relación meramente inductiva /empírica. Sino que es una relación que está en el medio, es más fuerte que la relación meramente inductiva y más débil que la conexión lógica, porque puede fallar en ocasiones (es «derrotable»/ *defeasible*). Cuando aprendemos el lenguaje, por ejemplo, la palabra «dolor» o «tigre» o «rojo», aprendemos en qué circunstancias estamos justificados a emitir una serie amplia y en principio indeterminada de oraciones en las que ocurre la palabra bajo consideración. Este conjunto de oraciones contribuye a la determinación del significado de la palabra en cuestión. Pero este aprendizaje es posible bajo dos supuestos. Por un lado, se suponen ciertas condiciones de normalidad en el mundo, aprendemos el uso del lenguaje ante objetos paradigmáticos, no ante casos límite. Por otra parte, también se supone un acuerdo entre los hablantes, se supone un acuerdo acerca de cómo el mundo es, la aceptación común de una serie de «verdades». (Z 350-352)

Antes de pasar a la evaluación de los argumentos acerca del problema mente cuerpo, quisiera detenerme a recordar algunas de las cuestiones acerca de los conceptos mentales y sus criterios de aplicación que Wittgenstein señaló y que serán de suma importancia para la evaluación de los argumentos que voy a mencionar luego.

1. Los conceptos mentales, en el caso de la tercera persona, se aplican de acuerdo con criterios conductuales (tanto la conducta manifiesta como las expresiones faciales).
2. Los conceptos mentales, en el caso de la autoatribución, de la atribución en primera persona, se aplican sin criterio, ie. se supone que uno se autoatribuye estados mentales sin apelar a base evidencial alguna, es decir que no hay ningún estado previo ni situación previa que funcione como razón para este tipo de atribuciones (obviamente, siguiendo las ideas wittgensteinianas esto no impide que haya causas de estos hechos). (PI 289, Z 659)
3. A pesar de las diferencias de criterios de atribución para los dos casos anteriores, Wittgenstein se encarga de señalar que esto no significa que haya dos conceptos mentales, sino uno sólo cuyos criterios varían cuando se aplican a la primera o la tercera persona. (Z 472) Este

---

7 Keil (1989) testea empíricamente la propuesta de clases naturales desarrollada en Quine (1967).

es un rasgo peculiar de los conceptos mentales, y, en parte, no darse cuenta de esta peculiaridad es lo que lleva a confusiones conceptuales que sería bueno remediar. (Z 220, 221, 472 PI 377)

4. Los criterios de atribución son públicos, y además por ser los criterios que constituyen el concepto, son suficientes para la maestría de los conceptos mentales; «lo que está en la caja» es irrelevante para el correcto uso de los conceptos mentales con propósitos comunicativos. (PI # 293) Tomando la distinción entre los dos componentes de los conceptos de clase natural, esto supone que «el escarabajo de la caja» funciona como «esencia» del concepto, y su correcta descripción en términos científicos es irrelevante para el uso del concepto. Sin embargo, si aceptamos que se trata de conceptos de clase natural, esta tendencia a buscar estas redescripciones fisiológicas o funcionales de lo que hay en la caja, son parte del significado del término. En este punto mi propuesta se aleja de la de Wittgenstein. Lo importante es que Wittgenstein estuvo en lo correcto en lo que hace al uso del lenguaje con propósitos comunicativos. Esto es muy importante porque objeciones a ciertas versiones de respuesta al problema mente-cuerpo, como por ejemplo al funcionalismo analítico por su carácter holista, que sostiene que los estados mentales no pueden ser funcionales (no pueden tener una esencia funcional) porque nadie puede conocer todas las conexiones y sin embargo usamos correctamente los términos, o la que sostiene en contra de la teoría de la identidad que los estados mentales no pueden ser estados neurofisiológicos porque no chequeemos que un cierto estado neurofisiológico se dé para afirmar que se da un cierto estado mental, confunden el plano criterial comunicativo de los conceptos mentales con el plano «esencial», o científico de redescripción de los fenómenos señalados por dichos conceptos.
5. Hay al menos algunos conceptos mentales que son conceptos reconocitivos: suponen la capacidad efectiva de discriminar ciertos estados del mundo. Los conceptos mentales, como los conceptos de color, tienen la peculiaridad de que la maestría del concepto presupone haber estado en ciertas situaciones paradigmáticas, y tener ciertas capacidades fisiológicas funcionando correctamente. (Z 267: «¿Se supone que es un hecho empírico que quien haya tenido una experiencia pueda imaginársela, y que otra persona *no* pueda? (¿Cómo sé que el ciego es capaz de imaginarse los colores?) Sin embargo, no puede jugar determinado juego de lenguaje (no puede aprenderlo). Pero, ¿Cómo? ¿Es esto de orden empírico, o bien es así *eo ipso*? Es lo último.») Si alguien afirma «Esto es rojo» y se le pregunta por qué afirma esto, la respuesta es «porque sé castellano». De la misma manera, si alguien afirma saber qué es tener un dolor de cabeza, pero al mismo tiempo afirma que jamás tuvo dolor de cabeza, dudaríamos de que realmente esté en posesión del concepto «dolor de cabeza.»
6. La gramática de los conceptos mentales debe estar dada antes de la adquisición de los términos mentalistas. Todo concepto supone la existencia de un acuerdo previo, de condiciones normales de adquisición tales que, si son violadas, los conceptos dejan de tener aplicación. (Z 350-2)
7. ¿Cómo se aprende el juego de lenguaje de las sensaciones? Por reemplazo de las expresiones primitivas de sensación por el lenguaje de las sensaciones. (PI 244). Más allá de que la explicación de la adquisición de los conceptos mentalistas que propone Wittgenstein sea la adecuada, incluir este tema es importante porque permite especificar las condiciones normales de aplicación de los conceptos, que son las que permitieron el aprendizaje.
8. Tampoco es necesario aceptar la teoría expresivista de los enunciados en primera persona del singular que Wittgenstein sostuvo. Como muestra Hacker 1972 no se sigue necesariamente

de 7. Independientemente de que uno esté dispuesto a aceptar la teoría que Wittgenstein expone acerca del aprendizaje de ciertos conceptos mentalistas y su teoría de los *avowals* (en lo personal, yo no suscribiría ninguna de las dos, pero argumentar esto queda fuera del alcance de este trabajo), ambos conjuntos de consideraciones aportan más indicios para sostener que hay una asimetría gramatical entre la primera persona del singular y las demás.

9. Algunas peculiaridades de lo mental, pueden ser explicadas como hechos gramaticales del lenguaje mentalista:
  - 9.1. La autoridad de la primera persona. (PI 247)
  - 9.2. La privacidad del ámbito mental. (PI 248)
10. Hay una marcada heterogeneidad de los conceptos mentales. (plan para el tratamiento de los conceptos psicológicos: no son todos reducibles a un único tipo Z 472) No todos los conceptos mentales tienen el mismo comportamiento gramatical.

### Algunos argumentos clásicos en torno del problema mente-cuerpo

En este apartado voy a examinar tres argumentos famosos en las discusiones mente-cuerpo contemporáneas, los dos primeros responsables, a mi juicio, del desplazamiento hacia el dualismo en las discusiones más recientes en estos ámbitos, el tercero, el que lleva al eliminativismo. Voy a tratar de mostrar que ninguno de estos argumentos funciona si los conceptos mentales se entienden de la manera en que yo propongo, por lo que alguna versión de la teoría de la identidad y/o del funcionalismo podrían estar en lo correcto como respuestas a la pregunta metafísica acerca de la naturaleza de los estados mentales.

#### *El argumento de Kripke en contra de la teoría de la identidad psicofísica*

Los argumentos de Kripke fueron propuestos originariamente en contra de la versión dominante de fisicalismo en la década de los '60, la teoría de la identidad psicofísica, que sostenía que ciertos estados mentales, específicamente las sensaciones, son contingentemente idénticas a estados cerebrales. Los argumentos de Kripke están basados en los siguientes supuestos.

- 1) Si una identidad entre dos designadores rígidos<sup>8</sup> es verdadera, entonces es necesaria.
- 2) Algo es necesario si no hubiera podido ser de otra manera que como es.
- 3) Hay necesidades *a posteriori*, esto es hay cosas que no podrían haber sido de otra manera que como son, pero que descubrimos que son como son por investigación empírica. Tal es el caso de las esencias de los términos de clase natural, que son necesarias, pero se descubren por investigación científica. (p. 110)<sup>9</sup>
- 4) Ciertas identidades (por ej. «Calor» = «energía cinética media de las moléculas») nos producen una «ilusión de contingencia» que es necesario explicar (porque creemos que podría haber calor independientemente de la energía cinética de las moléculas, o que el calor es pro-

8 Una expresión designa rígidamente si designa la misma entidad en todo mundo posible.

9 También hay enunciados contingentes *a priori*. Por ejemplo, «calor = aquello que se siente a través de sensaciones S», es *a priori* en la medida en que la descripción involucrada es usada para fijar la referencia de calor, y por lo tanto no es necesario averiguar cómo el mundo es para conocer su verdad, pero es contingente porque dicha propiedad está sólo contingentemente atada al calor, seres con otro aparato perceptivo podrían tener diferentes sensaciones.



ducido, por decir algo, por moléculas en estado de reposo, pero esta creencia es ilusoria en la medida en que toda identidad es necesaria). La ilusión de contingencia depende del hecho de que para fijar la referencia de los términos, en el caso de algunos términos de clase natural —por ejemplo «calor»— usamos descripciones tales como «aquello que nos produce la sensación de calor». Estas descripciones no son idénticas en significado a los términos cuya referencia fijan, pero especifican la evidencia en base a la cual estamos autorizados a utilizar un cierto término en nuestro hablar cotidiano. La ilusión de contingencia se explica, entonces, por el hecho de que esta base evidencial sólo determina contingentemente la naturaleza de los términos. Es decir, lo denotado efectivamente por el término «calor» es (necesariamente) «energía cinética media de las moléculas.» Pero nuestro mundo podría haber sido distinto, en el sentido de que lo que nos produce la sensación de calor podría haber sido la luz (para usar el ejemplo de Kripke). Y esta contingencia explica la ilusión de contingencia que se produce en el caso del enunciado necesario «calor = energía cinética media de las moléculas». Hay un hiato (contingente) entre la esencia del término de clase natural y aquellas manifestaciones de esa esencia que dependen del hecho de que tengamos el aparato perceptivo que tenemos; esto nos hace pensar que la esencia era contingente, cuando lo contingente son nuestras capacidades cognitivas y perceptivas.

El primer argumento de Kripke en contra de la TI se basa en una cierta asimetría que él encuentra entre el caso psicofísico y los casos mencionados. La idea, básicamente, es que la pretendida ilusión de contingencia del caso psicofísico, no puede ser explicada de la manera en que es exitosamente explicada esta ilusión en el caso del calor o del agua. El argumento podría, entonces, esquematizarse así:

- (i) Podemos **concebir** la estimulación de las fibras-C sin que sean sentidas como dolor.
- (ii) Si **es posible** la estimulación de las fibras C sin el dolor, entonces no hay una relación necesaria entre ambos.
- (iii) Toda identidad, si es verdadera, es necesaria.
- (iv) No hay una identidad necesaria entre dolor y estimulación de las fibras C.
- (v) Por lo tanto, la teoría de la identidad es falsa.

El paso de (i) al antecedente de (ii) para hacer el *modus ponens* necesario para llegar a la conclusión es impedido, en el caso del calor, porque la contingencia es una ilusión, esto es podemos afirmar que es concebible la situación en la que el calor no es la energía cinética media de las moléculas aunque esta situación no es posible. El punto de Kripke es que en el caso del dolor la mera concebibilidad implica la posibilidad, porque no hay manera de mostrar que tal situación es concebible en base a una ilusión que, en el caso del calor, depende del hecho de que lo que conocemos/percibimos del calor (la sensación de calor) no es el calor mismo (sino aquello que es producido/causado por el calor), en tanto que en el caso del dolor lo que conocemos es la sensación de dolor misma, el dolor *es* la sensación de dolor. No hay una distancia entre la base evidencial en virtud de la cual atribuimos dolor y el dolor mismo. En palabras de Kripke:

«Alguien puede estar en la misma situación epistémica en la que estaría si hubiera calor, aún en ausencia de calor, simplemente sintiendo la sensación de calor; y aún en la presencia de calor, él puede tener la misma evidencia que tendría en ausencia de calor, simplemente



careciendo de la sensación S. Ninguna posibilidad similar existe en el caso del dolor y otros fenómenos mentales. **Estar en la misma situación epistémica que se daría si uno tuviera dolor es tener un dolor; estar en la misma situación epistémica que se daría en la ausencia de dolor es no tener dolor.** La aparente contingencia de la conexión entre el estado mental y el correspondiente estado cerebral no puede ser explicada por alguna suerte de análogo cualitativo como en el caso del calor.» (p. 152, sus itálicas, mis negritas).

El corazón del argumento kripkeano consiste, entonces, en mostrar que la ilusión de contingencia no puede ser explicada en el caso de los fenómenos concientes. La respuesta fisicalista a este argumento deberá consistir, por lo tanto, en una explicación de esta ilusión de contingencia. García Carpintero (2000), adoptando ideas de Loar (1997) nos muestra el camino para explicar la ilusión de contingencia presente en el caso de «dolor = estimulación de las fibras-C». Básicamente, el camino que proponen depende de mostrar que hay dos modos de presentación del estado cualitativo (que, si los teóricos de la identidad están en lo correcto, será un estado del cerebro), uno teórico y otro por contacto, o reconocitivo. El teórico es aquel que incluye la ubicación del estado cualitativo en un espacio cualitativo más amplio, esto es el conjunto de relaciones entre ese estado cualitativo y otros. (si se trata de colores, relaciones de similaridad entre colores, por ejemplo que el naranja es más semejante al rojo que al verde). El otro, consiste en una cierta capacidad de discriminación que sólo es posible adquirir una vez que se ha estado en el estado en cuestión (por eso se opone este modo de presentación al «teórico», que podría tener alguien sin haber estado en el estado descrito). Al separar los modos de presentación o conceptos de lo referido por esos conceptos, es posible sostener que dos modos de presentación diferentes refieren al mismo estado en el mundo, pero que dado que es posible poseer un concepto/modo de presentación sin el otro, entonces podemos entender por qué es concebible que se dé una cosa sin la otra (en nuestro caso, la estimulación de las fibras C sin el dolor, y viceversa). Básicamente, entonces, la mejor estrategia fisicalista para evitar el dualismo es separar el plano conceptual del plano ontológico, planos que Kripke había unido en su argumento.

Siguiendo la propuesta presentada al comienzo de este trabajo, habría que sostener que poseer el concepto de «dolor» supone poder distinguir a un estado como estado de dolor, tanto desde la perspectiva de la primera persona como de la tercera. Los distintos modos de presentación mencionados en el argumento contra Kripke constituyen un único y unificado concepto ordinario de «dolor». Si esto es así, no es correcta la afirmación que resalté en negrita en la cita de Kripke. En efecto, a pesar de lo afirmado por Kripke, sentir un dolor no es lo mismo que estar en la situación epistémica que se daría si uno tuviera un dolor, porque ésta supone además de la sensación de dolor (que puede o no ser idéntica a un estado cerebral) la posesión de un cierto concepto, y toda vez que poseemos un concepto y lo aplicamos existe, al menos en principio, la posibilidad de que lo estemos aplicando incorrectamente, esto es que apliquemos el concepto «dolor» en ausencia de una genuina sensación de dolor<sup>10</sup>. El destino de la tesis fisicalista no puede depender exclusivamente de lo que nosotros creamos o digamos acerca de nosotros mismos.

Es importante señalar que, en el argumento de Kripke, está supuesta una comprensión del concepto «dolor» diferente a la que yo propuse en este trabajo. Las negritas que puse en la cita de

10 Nótese que este tipo de errores son muy frecuentes en los niños cuando todavía no dominan adecuadamente los conceptos de sensaciones; en esos casos suele ocurrir que afirmen «Me duele la panza» en lugar de «Tengo hambre» o «Tengo frío» en lugar de «Quiero ir al baño».

Kripke apuntan a mostrar que estas afirmaciones están hechas teniendo en cuenta exclusivamente la perspectiva de la primera persona para la consideración del manejo del concepto «dolor», ya que, si tomamos en cuenta la heteroatribución de «dolor» parece razonable pensar que la evidencia necesaria para este tipo de adscripciones no incluye que uno mismo (el atributor) tenga dolor, ni dependen del acceso a cualidad fenomenológica alguna. Nuestros conceptos mentalistas ordinarios, sin duda, suponen esta dualidad de perspectivas, la de la primera y la de la tercera persona. Por el contrario, argumentos como el de Kripke sólo se aplican a un concepto de dolor que se agota en la primera persona. Así, creo que podríamos enfrentar a Kripke con el siguiente dilema: o bien en su argumento usa «dolor» en el sentido que habitualmente tiene en nuestro lenguaje ordinario, en cuyo caso su afirmación es falsa, porque para el caso de la atribución de dolor en la tercera persona la evidencia disponible no se agota en sentir el dolor, o bien está usando «dolor» en un sentido un poco diferente del sentido ordinario del término.

*El argumento en contra de la funcionalización (zombis y espectros invertidos)*

Básicamente, el funcionalismo analítico sostiene que si podemos ofrecer un análisis funcional de un concepto dado, por ej. «dolor» entonces, tal como mostró en varios trabajos D. Lewis, por un procedimiento reductivo similar al de la Oración de Ramsey es posible sostener una posición fisicalista de acuerdo con la cual, todo queda fijado una vez que queda fijado el mundo físico: todo lo demás son meras maneras alternativas de hablar, que una vez traducidas en términos adecuados, no nombran en el mundo entidades diferentes de las nombradas por la física, sino sólo redescubren el mundo físico. El paso fundamental en esta línea argumentativa en contra de esta versión del fisicalismo (y en favor del dualismo) es mostrar que no es posible ofrecer un análisis funcional adecuado de ciertos conceptos mentales. Es en este punto donde se presentan dos argumentos claves: el argumento de los zombis y el del espectro invertido (en sus diversas versiones). El núcleo de estos argumentos podría reconstruirse en los siguientes términos:

- (i) Hay experiencia conciente (fenoménica) en nuestro mundo. (En mis palabras: Hay descripciones verdaderas de nuestro mundo que involucran términos del lenguaje ordinario tales como «sensación de dolor», «parecerme rojo», etc.)
- (ii) Podemos concebir un escenario físicamente/funcionalmente indistinguible de nuestro mundo donde los conceptos fenoménicos son aplicados con verdad de una manera diferente a como son aplicados en el nuestro. Esto es, podemos imaginar una situación donde un individuo se comporte en todos los aspectos exactamente igual que un individuo dado en el mundo actual, pero donde tal individuo no posea estado cualitativo alguno (o posea estados cualitativos diferentes); por ejemplo que un individuo exhiba todas las conductas asociadas al dolor en nuestro mundo, que se encuentre en el estado neurofisiológico en el que se encuentran los individuos de nuestro mundo cuando sienten dolor, pero que no sienta dolor en lo absoluto (argumento del zombi) o que sienta otra cosa, por ejemplo una leve comezón (argumento del espectro invertido).
- (iii) Si algo es concebible, entonces es posible.
- (iv) Si la inversión del espectro (o *mutatis mutandis* el mundo zombi) es posible, entonces la superveniencia lógica entre lo físico y lo cualitativo falla.
- (v) El fisicalismo es verdadero si y sólo si todos los hechos positivos son implicados por los hechos físicos, i.e. si la superveniencia lógica se sostiene. En otras palabras, **alguien que**

**conoce todos los hechos físicos y posee los conceptos de nivel superior, también conoce los hechos de nivel superior<sup>11</sup>.**

Por lo tanto, (vi) El fisicalismo es falso.

Nótese que este argumento, una vez más, funciona sólo para una cierta lectura de la premisa (ii). Si «concepto fenoménico» en la premisa (ii) se entiende como concepto mentalista *ordinario*, por ej. tener la sensación de rojo, entonces la premisa (ii) es falsa. Nuestros conceptos mentalistas ordinarios, como traté de explicitar en el primer apartado, suponen criterios diferentes de aplicación para la primera y la tercera persona de un *mismo concepto*. Nuestros conceptos ordinarios incluyen un elemento reconocitivo tal que si uno no ha pasado por las experiencias adecuadas no posee el concepto en cuestión. Un zombi no tendría maestría genuina en el uso de los conceptos mentalistas ordinarios, tal como ocurre en el caso del ciego y los conceptos de colores<sup>12</sup>. Sólo es verdadera la premisa (ii) si por «concepto fenoménico» entendemos algo como una etiqueta (y sólo eso) del estado cualitativo en cuestión. Chalmers ha defendido la idea de que cuando pensamos acerca de experiencias como las de tener la sensación de rojo hay escondidos en realidad cuatro conceptos fenoménicos diferentes. (1) El concepto relacional comunitario («la cualidad causada en los observadores normales de mi comunidad por las cosas rojas»), (2) el concepto relacional individual («la cualidad normalmente causada en mí por cosas rojas»), (3) el concepto indexical («la cualidad que estoy experimentando ahora») y (4) el concepto cualitativo (R)<sup>13</sup>. Chalmers se preocupa explícitamente por señalar que si bien se trata de cuatro conceptos, la propiedad en el mundo que hace verdadera una afirmación que contiene a cualquiera de ellos cuatro es la misma esto es, es necesario *a posteriori* que (1) = (2) = (3) = (4). Mi punto es que todas las distinciones que Chalmers considera «distintos conceptos» son partes constitutivas del concepto mentalista ordinario. El argumento antifisicalista de los zombis sólo funciona para una concepción de los conceptos mentalistas diferente de la que propongo en este trabajo, una como la propuesta de Chalmers, de grano más fino que la que el lenguaje mentalista ordinario nos permite realizar.

### *El argumento eliminativista de Churchland*

Hace veinticinco años Paul Churchland escribió un artículo en el que defendió una respuesta al problema mente-cuerpo que él denominó «materialismo eliminativo». Básicamente, la estrategia de Churchland fue considerar a nuestra concepción de sentido común acerca de nuestras mentes como un tipo de «teoría», más precisamente como una teoría incorrecta que podría ser reemplazada por una mejor teoría (la neurofisiología madura) capaz de explicar los mismos fenómenos. La idea compartida por Churchland, Lewis y muchos otros es que la «*Folk Psychology*» es un conjunto de leyes

11 Estos argumentos parten de una específica caracterización del fisicalismo; la definición que proponen en la siguiente: «el materialismo (fisicalismo) es verdadero si todos los hechos positivos acerca del mundo son lógicamente supervenientes en forma global a los hechos físicos, ...o el materialismo es verdadero si todos los hechos positivos acerca de nuestro mundo están implicado por los hechos físicos» (Chalmers, 1996 p. 70-1).

12 Con una línea argumentativa similar, *mutatis mutandis*, se puede hacer frente al argumento antifisicalista de Mary de Jackson.

13 Nótese que la manera en que Chalmers caracteriza a los tres primeros incluye el giro «la cualidad que...» aunque podría reescribirse en forma neutra de la siguiente manera: «es aquel estado que ...», por su parte el cuarto es una mera etiqueta. La distinción entre los cuatro conceptos podría mantenerse aún sin comprometerse con los *qualia* que Chalmers (y otros) reconocen.

causales que se usan para explicar y predecir la conducta. Estas leyes conectan inputs perceptuales y otros con la conducta, postulando un tipo especial de entidades teóricas —los estados mentales— que tienen, dentro de esta teoría, el mismo estatus que las demás entidades teóricas en cualquier teoría científica, esto es son como «átomos»: entidades no-observables que se suponen con el fin de explicar los hechos observables.

El argumento en favor del materialismo eliminativo puede ser reconstruido en los siguientes términos:

- (i) La *Folk Psychology* es una teoría empírica.
  - (ii) Las teorías empíricas tienen que satisfacer los siguientes requisitos:
    1. Tienen que explicar todos los fenómenos bajo su dominio.
    2. Tienen que ser programas de investigación progresivos.
    3. Tienen que ser coherentes y continuas con otras teorías bien establecidas.
  - (iii) La *Folk Psychology* no satisface 1., 2. ni 3.
- por lo tanto, (iv) la *Folk Psychology* debe ser eliminada.

Creo que la hipótesis de entender a los conceptos mentales como conceptos de clase natural permite preservar el estatus teórico de la *Folk Psychology* de una manera fructífera, evitando su eliminación. De acuerdo con mi propuesta la *Folk Psychology* ha de ser entendida como una red de conceptos ordinarios, incluyendo por lo menos tres tipos de conceptos mentales: conceptos de sensación, de emoción y de actitud proposicional. Cada uno de estos conceptos tiene una «gramática», una teoría o cuerpo de conocimiento que incluye muchos otros conceptos mentales, algunos hechos conductuales (tanto lingüísticos como no-verbales), algunas manifestaciones faciales, etc. Esta teoría constituye nuestra *Folk Psychology*, el conjunto de componentes estereotípicos de nuestros conceptos psicológicos. Cuando aprendemos el significado de estos conceptos adquirimos maestría en esta teoría, y nos volvemos capaces de comunicar estados mentales a los demás y de explicar y predecir acciones propias y ajenas. En mi opinión, podemos distinguir al menos dos sentidos en los cuales podemos decir que la *Folk Psychology* es una teoría. Por un lado, podemos considerarla una teoría porque aceptamos la idea de que los estados mentales son estados postulados sobre la base de conducta observable, para explicar y predecir la conducta humana. Este es el sentido original en el que Lewis y Churchland afirmaron que la *Folk Psychology* es una teoría. Pero no es este sentido el que me interesa aquí. En efecto, hay otro sentido más amplio de teoría: es la idea de que cualquier conjunto más o menos coherente de creencias acerca de un tema es una teoría, y es este mismo sentido en el cual aceptamos que antes de conocer la microestructura del agua ya teníamos una «teoría del agua». En este segundo sentido quiero sostener que la *Folk Psychology* es una teoría. Pero este segundo sentido no nos compromete con la idea de que sólo podemos observar conducta. Por el contrario, es completamente compatible con la idea de que observamos directamente nuestros estados mentales y los ajenos, esto es que no «inferimos» los estados mentales de los demás sobre la base de la conducta o la expresión facial<sup>14</sup>.

14 En mi opinión, el pasaje frecuentemente citado de Wittgenstein (Z 225) en apoyo a la idea de que la *Folk Psychology* no puede ser una teoría sólo es adecuado para rechazar el primer sentido de teoría mencionado en este texto, no para el segundo. Este segundo permite aceptar la idea de que tenemos un cierto conocimiento común de nuestra vida mental, que de alguna manera constituye una teoría, aunque no completamente empírica; pero para alguien que como yo rechaza el dogma empirista de la distinción tajante entre lo analítico y lo sintético, si se trata de una teoría conceptual o empírica no es demasiado importante.

Pero hay también teorías científicas relacionadas con nuestros conceptos psicológicos cotidianos. De acuerdo con la explicación de los conceptos de clase natural que propuse arriba, los conceptos mentales también tienen un componente indexical. Este componente puede ser completado con más teoría, tales como la propuesta para descubrir la estructura oculta de estos conceptos mentales. Pero estas teorías son continuas con nuestras teorías folk, dado que son construídas con el fin de explicar más profundamente hechos que ya han sido explicados por la teoría que constituye la «gramática» de estos conceptos. Las teorías psicológicas de sentido común y científicas están (como consecuencia de mi punto de vista) en la misma relación que nuestra teoría folk acerca del agua y la química, relación que fundamenta la identidad agua = H<sub>2</sub>O.

De estas consideraciones se sigue además que la posición eliminativista es insostenible porque nuestra *Folk Psychology* es nuestra comprensión básica de nuestras vidas mentales, el punto inicial para la producción de investigación científica. Nuestro conocimiento cotidiano y científico forman un continuo, y nuestras teorías científicas pueden explicar algunos fenómenos de sentido común. Por supuesto, nuestra comprensión ordinaria puede cambiar a causa de descubrimientos científicos pero, en última instancia los mismos fenómenos que constituyen nuestra comprensión ordinaria del mundo y de nosotros mismos tienen que ser incluidos en la imagen científica del mundo porque no puede contradecirla. Como dije antes los científicos buscan la «esencia» de agua, cuando descubren que el agua es H<sub>2</sub>O no eliminan el agua, sino que la incorporan en la imagen científica del mundo. Si pretendemos que la psicología, al tratar con estados informacionales o neurofisiológicos pueda eventualmente cambiar y transformarse en algo donde no tengan lugar nociones como las de «creencia», «deseo», «dolor», «miedo» etc. entonces habrá habido un cambio de tema, y ya no estaríamos haciendo más psicología.

### **Conclusión. A mitad de camino entre el funcionalismo y la teoría de la identidad**

He intentado mostrar que ciertos argumentos famosos en contra del funcionalismo y la teoría de la identidad son inconclusivos, así como el argumento pro-eliminativista de Churchland. Queda, por lo tanto, un espacio abierto para defender la idea de que las «esencias» de los estados mentales podrán ser descritas por una psicología/neurofisiología/antropología/sociología/etc. madura en sus propios términos, en un futuro tal vez no tan lejano. Sostengo que entendiendo a los conceptos mentales como conceptos de clase natural se puede proyectar un porvenir optimista para la psicología científica y más en general, para una comprensión naturalista más profunda de nuestra vida mental. Entiendo asimismo que esta propuesta hace posible delinear una respuesta al problema mente-cuerpo combinando aspectos del funcionalismo y de la teoría de la identidad, compatibles con los avances recientes en psicología cognitiva y neurociencias: en la medida en que se logre identificar empíricamente una «esencia» neurofisiológica para ciertos fenómenos mentales, la teoría de la identidad habrá triunfado para ese subconjunto de fenómenos mentales, en la medida en que las «esencias» resulten ser funcionales, el psicofuncionalismo habrá dado en el clavo. La heterogeneidad remarcada en los estereotipos mentalistas nos permiten visualizar un futuro heterogéneo e interdisciplinario para dar cuenta en forma acabada de nuestra compleja y riquísima vida mental.

### **Referencias**

- Chalmers, D. (1996) *The Conscious Mind*. Oxford University Press.  
 Chalmers, D. (2003) «The Content and Epistemology of Phenomenal Beliefs», in Smith, Q. and Jokić, A. *Aspects of Consciousness*, Oxford University Press.

- Churchland, P. (1981) «Eliminative Materialism and the Propositional Attitudes», in Churchland P. (1989) *A Neurocomputational Perspective*, Cambridge, MIT Press.
- García Carpintero, M (2000) «Razones para el dualismo», en Chacón, y otros, *Pensar la mente*, Biblioteca nueva, Madrid.
- Goldman, A. (1993) «The psychology of Folk psychology», *BBS* 16.
- Greenwood, J. (1991) *The Future of Folk psychology*, Cambridge University Press.
- Griffiths, P. (1997) *What emotions really are*, Chicago University Press, Chicago.
- Hacker, P.M.S. (1972/1986): *Insight and Illusion*, Clarendon Press, Oxford.
- Keil, F. (1989) *Concepts, Kinds and Cognitive Development*, Cambridge, MIT Press.
- Kripke, S. (1972): *Naming and Necessity*, Harvard University Press, Cambridge.
- Lewis, D. (1966) «An Argument for the Identity Theory», *Journal of Philosophy* 63, 17-25.
- Lewis, D. (1972) «Psychophysical and Theoretical Identifications», *Australasian Journal of Philosophy*, 50, 249-58.
- Loar, B. (1997) «Phenomenal states» en Block, N., Flanagan, O. y Guzeldere, G. (eds.) *The Nature of consciousness*, MIT Press, Cambridge.
- McGinn, C. (1991) «The Hidden Structure of Consciousness», in Mental concepts Ginn (1991) *The problem of consciousness*, Blackwell, Oxford.
- McGinn, C. (1993) *Problems in Philosophy*, Blackwell, Oxford.
- Malt, B. (1994) «Water is not H<sub>2</sub>O», *Cognitive Psychology* 27, 41-70.
- Margolis, E. and Laurence, S. (1998) *Concepts. Core Readings*, MIT Press.
- Medin, D. (1989) «Concepts and Conceptual Structure», *American Psychologist*, vol. 44, no. 12.
- Murphy, G. and Medin, D. (1985) «The rol of theories in conceptual coherence», reprinted in Laurence and Margolis (1998).
- Nagel, T. (1986) *The view from nowhere*. Oxford University Press.
- Putnam, H. (1975) «The meaning of 'meaning'» in Putnam (1975) *Mind, Language and Reality. Philosophical Papers. Vol 2*. Cambridge Univ Press.
- Quine, W.V.O. (1953) «Two dogmas of empiricism» in Quine, W.V.O. (1964) *From a Logical Point of View*, Harvard University Press, Harvard.
- Quine, W.V.O. (1967) «Natural Kinds» en Quine, W.V.O. (1969) *Ontological Relativity and other essays*, Columbia University Press, New York.
- Wittgenstein, L. (1953) *Philosophical Investigations*, Basil Blackwell, Oxford.
- Wittgenstein, L. (1967) *Zettel*, Basil Blackwell, Oxford.